MIDAS: An Agent Based Data Transcoding Services Framework

Rahul Ramachandran, Sunil Movva, Xiang Li, Sara Graves
Information Technology and Systems Center
University of Alabama in Huntsville
rramachandran@itsc.uah.edu

Earth Science Technology Conference, Paolo Alto, CA

June 22-25,2004

Data Preparation/Preprocessing

- Important step before analysis or mining
- Several preprocessing steps are general
  - Subsetting
  - Data Format Translation
  - Reprojection
- Subsetting: a data reduction step that helps in making the size of the data managable
- Data Format Translation: a format conversion step that allows the data to be stored in a familiar format

Project Objective

- To provide an *intelligent, automated data preprocessing* for Earth Science data

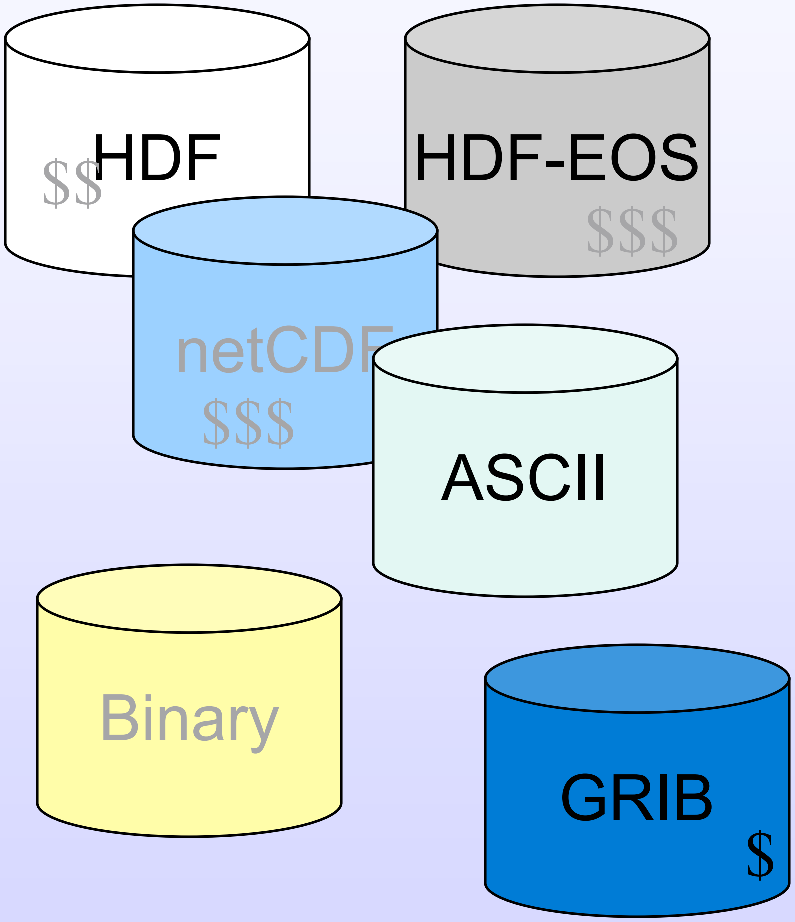
Challenges

- Requires a rich set of metadata
  - Structural metadata: to provide full description of the data file in bits & bytes, to allow application to read the data
  - Semantic metadata: to provide meaning of the data along with a context, to allow application to understand what it has read and how to use it
- Requires a Data Preprocessing Framework
  - Framework properties: Loosely coupled, asynchronous, intelligent and distributed

Metadata Solution

- Use Earth Science Markup Language (ESML) for structural metadata description
- Enhance ESML description by providing semantic metadata by leveraging ontologies

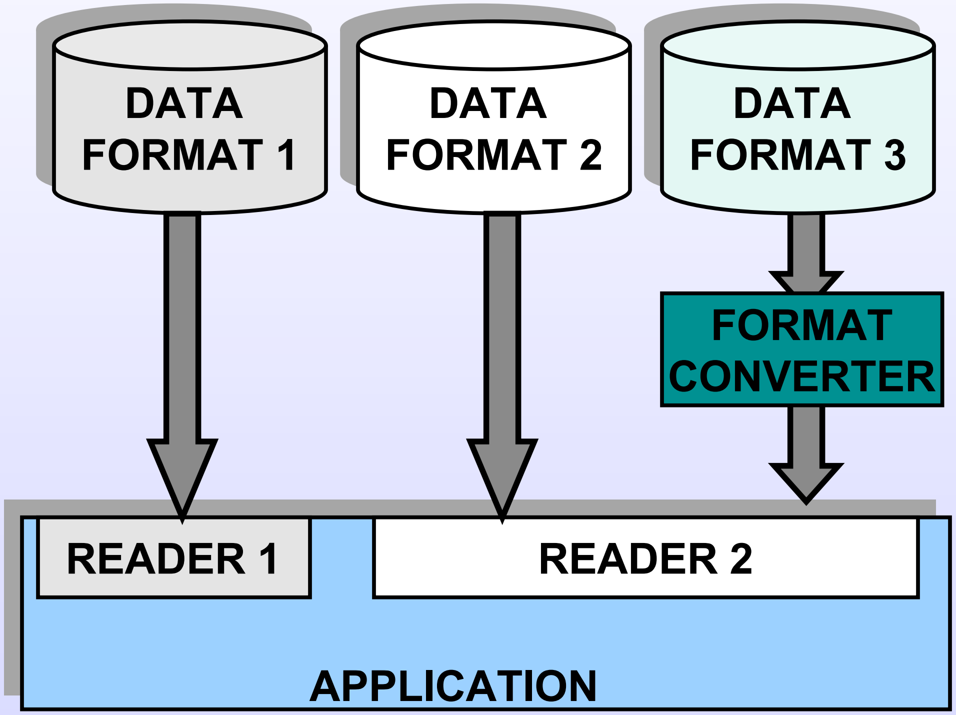
Background:
Earth Science Data Characteristics



- Different formats, types and structures (18 and counting for Atmospheric Science alone!)
- Some formats lack metadata where as others are metadata rich ($)

- **Heterogeneity leads to Data usability problem**

Background:
Data Usability Problem

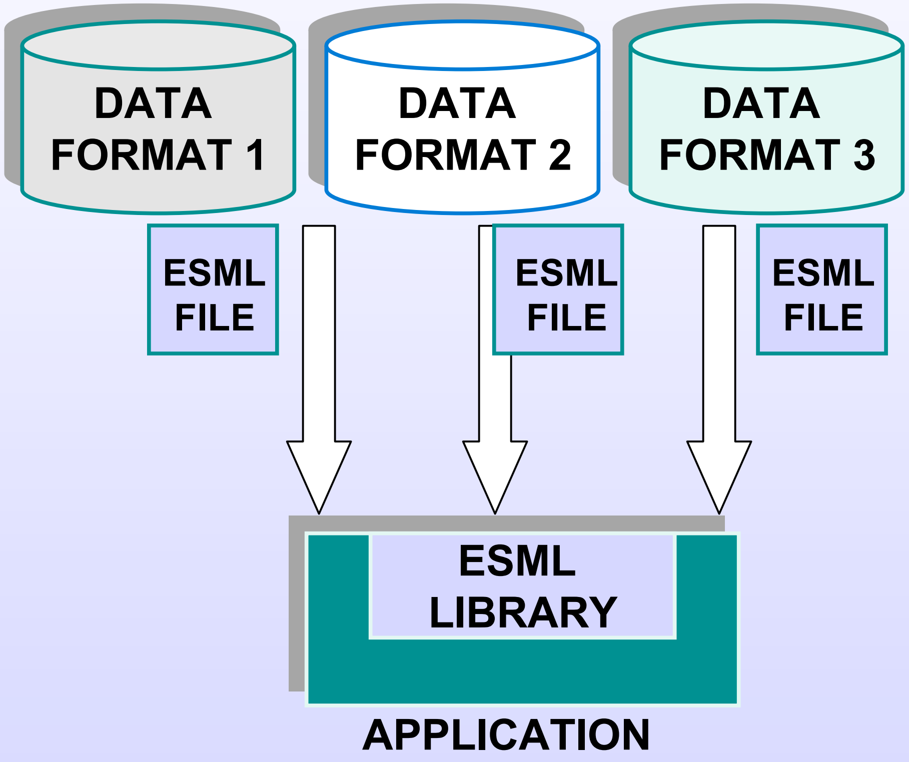


- Requires specialized code for every format
  - Difficult to assimilate new data types
  - Makes applications tightly coupled to data
- One possible solution - enforce a Standard Data Format
  - Not practical for legacy datasets

UAH
The University of Alabama in Huntsville

June 22-25,2004

Background:
Earth Science Markup Language (ESML)

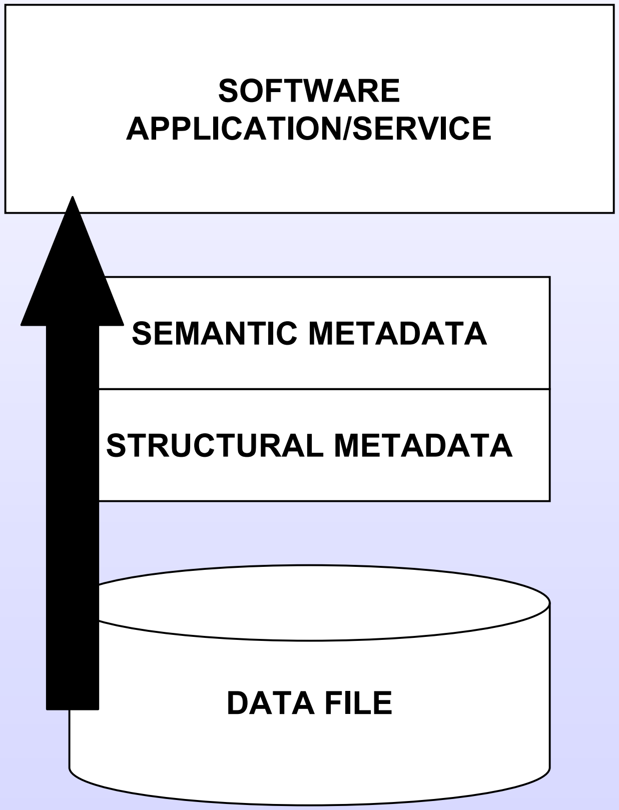


- ESML (external metadata) files containing the structural description of the data format
- Applications utilize these descriptions to figure out how to read the data files resulting in data interoperability for applications

UAH
The University of Alabama in Huntsville

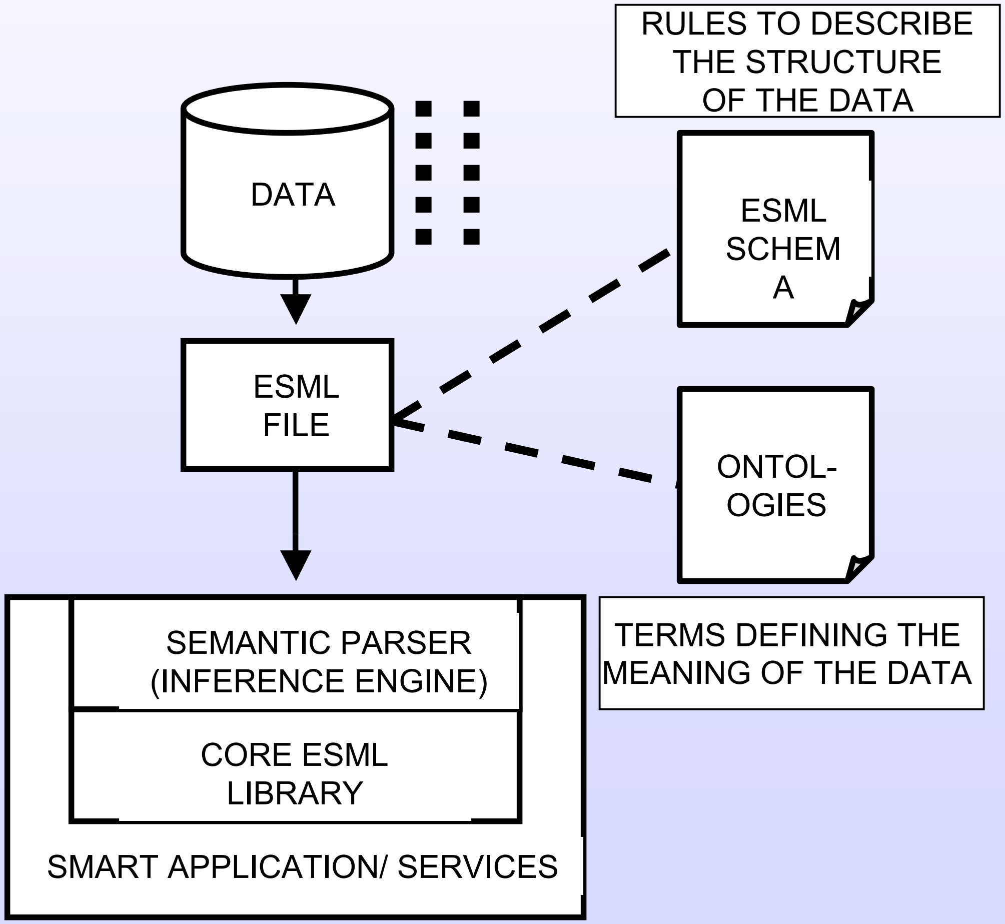
June 22-25,2004

Requirement: Vertical Metadata Integration



- Horizontal Metadata Integration
  - Mediation services
  - Yellow page services
- Vertical Metadata Integration focuses on semantics for the use of the data by an application
- Both Structural and Semantic metadata are required

Solution:
Extending ESML with Ontologies


RULES TO DESCRIBE THE STRUCTURE OF THE DATA
DATA
ESML SCHEM A
ESML FILE
ONTOL- OGIES
SEMANTIC PARSER (INFERENCE ENGINE)
CORE ESML LIBRARY
SMART APPLICATION/ SERVICES
TERMS DEFINING THE MEANING OF THE DATA

- ESML Schema provides structural metadata
- Extend ESML schema by embedding semantic terms in the ESML Description File to provide a complete description of the data
- Allow various science communities can create their own ontologies (for example, SWEET) and use them with ESML Description Files for their data

June 22-25,2004

UAH
The University of Alabama in Huntsville

Example:
Embedding Semantics in ESML

```
<<a:ESML xmlns:a="ESML" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="ESML.xsd" >
 <SyntacticMetaData>
  <Binary>
   <Structure instances="1" name="SampleSet">
    <Array occurs="100">
     <Array occurs="100">
      <Field name="UWind" type="Int32" order="LittleEndian"/>
     </Array>
    </Array>
    <Array occurs="100">
     <Array occurs="100">
      <Field name="DimX" type="Int32" order="LittleEndian"/>
     </Array>
    </Array>
    <Array occurs="100">
     <Array occurs="100">
      <Field name="DimY" type="Int32" order="LittleEndian"/>
     </Array>
    </Array>
   </Structure>
  </Binary>
 </SyntacticMetaData>
</a:ESML>
```

Original ESML Description File containing only structural metadata

ESML Description File embedded with Semantic Tags defined in separate ontologies

```
<a:ESML xmlns:a="ESML" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="ESML"
xmlns:daml="http://www.daml.org/2001/03/daml+oil#"
xmlns="http://www.itsc.uah.edu/esml-ex#">
   <SemanticMetaData>
       <Latitude rdf:ID="DimX"/>
       <Longitude rdf:ID="DimY"/>
       <DataField rdf:ID="UWind"/>
       <DataSet rdf:ID="SampleSet">
           <hasField rdf:resource="#DimX"/>
           <hasField rdf:resource="#DimY"/>
           <hasField rdf:resource="#UWind"/>
       </DataSet>
   </SemanticMetaData>
   <SyntacticMetaData>
       <Binary>
           <Structure instances="1" name="SampleSet">

           </Structure>
       </Binary>
   </SyntacticMetaData>
</a:ESML>
```

UAH

The University of Alabama in Huntsville

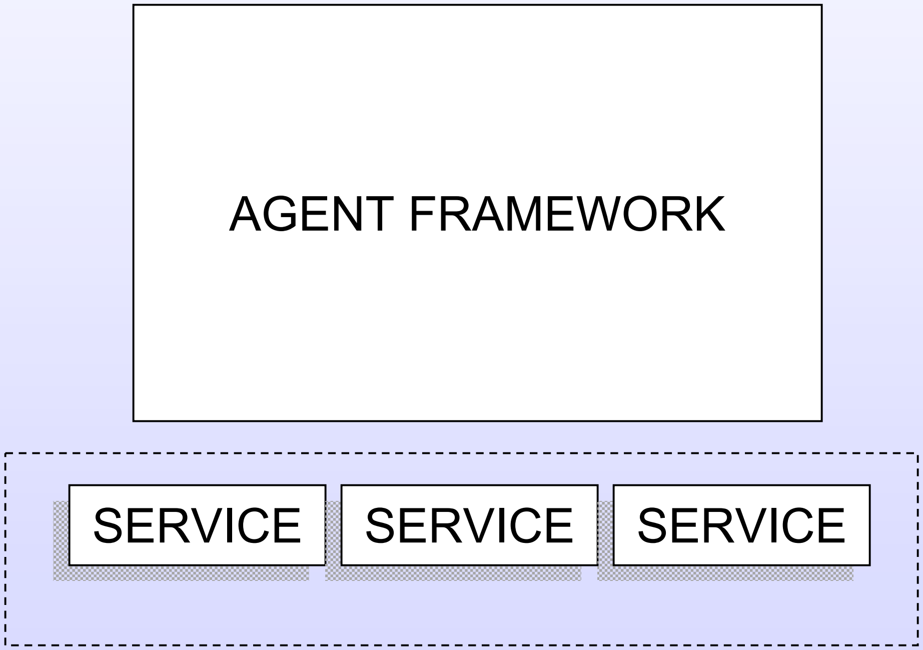
June 22-25,2004

Agent Framework Design: Layer Architecture

- *Infrastructure Layer*: provides the environment that agents can act upon, i.e. services
- *Agent Layer*: contains the agents used to achieve the overall goal of the framework
- *Organization Layer*: defines the organizational structure of the system which is important for agents interaction
- *Coordination Layer*: defines coordination methods required to resolve conflicts and select the next agent
- *Constraint Layer*: verifies whether the system goals are met and interfaces with the users/user interface

Agent Framework Design: Infrastructure Layer

- Consists of "transcoding" services
  - Search
  - Subsetting
  - Data Format Translation
  - Calibration
  - Navigation
  - Reprojection
  - Visualization
  - Aggregation
  - Fusion
  - Mining



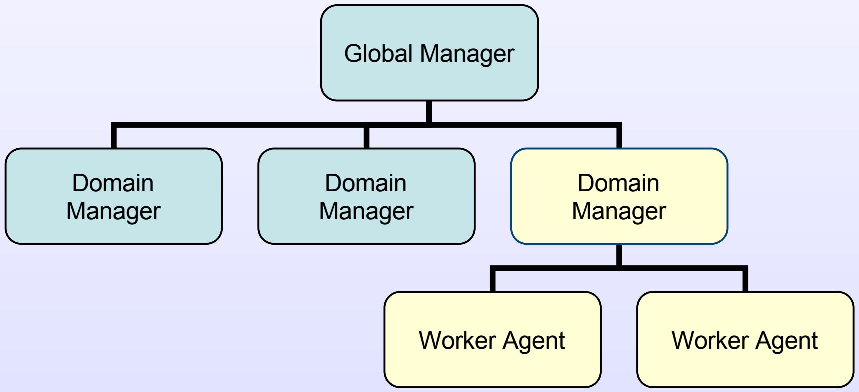
Agent Framework Design: Agent Layer (1)

- All the **agents** have the following features:
  - Role
  - Behavior- methods/functions that it can act upon
  - State – store information of itself and the world it perceives
  - Intelligence/Knowledge– to be able to make decisions via ontologies and a reasoning engine or via a machine learning algorithm or via heuristic algorithm
  - Communication protocol – to interact with other agents
- Types of Agents
  - Manager Agents
    - Global Manager
    - Domain Manager
  - Worker Agents

Agent Framework Design: Agent Layer (2)

- **Global Manager Agent:** Given an user input request, distribute the work and collate results
- **Domain Manager Agent:**
  - Keeps a registry of all the Worker agents in its domain
  - All Worker agents advertise their capabilities to the Manager Agent
  - Parses the incoming message and uses an ontology to find the “correct” Worker agent
  - Polls Worker agents for results
  - Fires and Hires Agents
- **Worker Agent**: Uses ESML semantic metadata and ontologies to **map** input message to API requirements of the Service
  - Example:
    - Navigate the data
    - Map Parameter Concept to Field Name(s)
    - Map Spatial Concept to Bounding Box
    - Map Temporal Concepts to Time Range

Agent Framework Design: Organization Layer

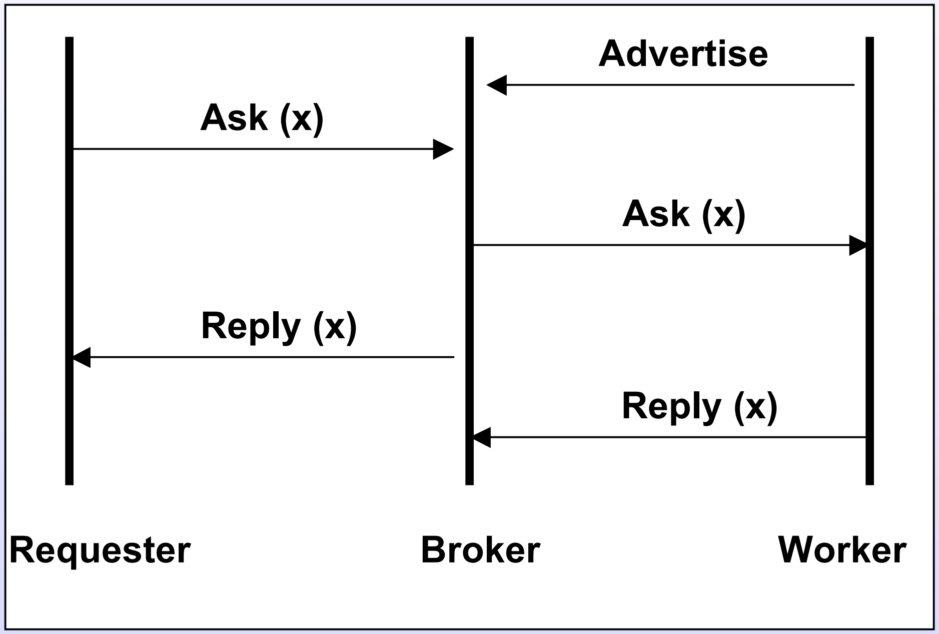


- A simple tree structure with a global manager and number of domains is used.
- Each of the domains contains a Manager agent and Worker agents
- Advantage: scalable design that will allow addition of new domains to the overall framework

June 22-25,2004

UAH
The University of Alabama in Huntsville

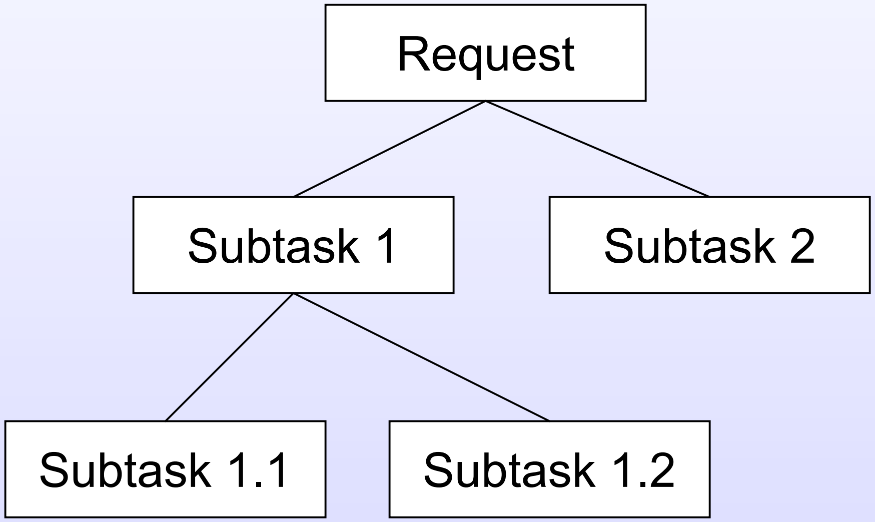
Agent Framework Design: Coordination Layer



- A broker model is used
- Advantage: unlike matchmaker or a contract-net, this model allows the broker to shoulder some responsibility of finding the right agent and returning the result

Agent Framework Design: Constraint Layer Design

- By using a tree to represents requests, one can check the goal achievement.
- When all the leaf and intermediary nodes are satisfied, resulting in completion of the root node, the task has been accomplished.

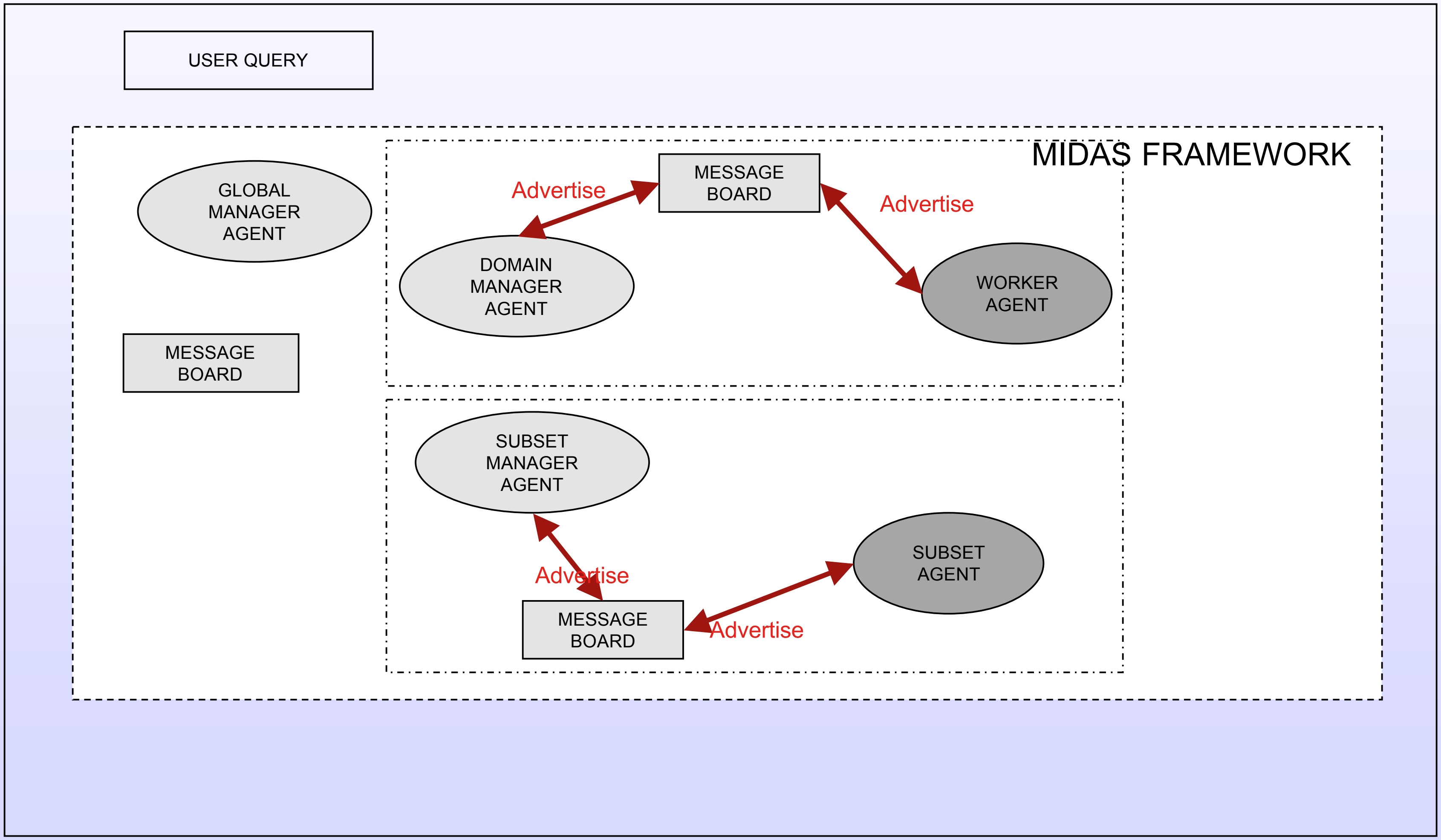


Agent Framework Design: Performatives

- Performatives are the permissible "speech acts" agents use to interact
- Partial set derived from KQML (Knowledge Query and Manipulation Language-UMBC)
- Basic Responses:
  - Error, Sorry
- Query:
  - Evaluate, AskStatus
- Capability:
  - Advertise

```
advertise
:content <performative>
:language KQML
:ontology <word>
:force <word>
:sender <word>
:receiver <word>
```

MIDAS in Action: Initialization



USER QUERY

MIDAS FRAMEWORK

GLOBAL MANAGER AGENT

MESSAGE BOARD

DOMAIN MANAGER AGENT

Advertise

MESSAGE BOARD

Advertise

WORKER AGENT

SUBSET MANAGER AGENT

Advertise

MESSAGE BOARD

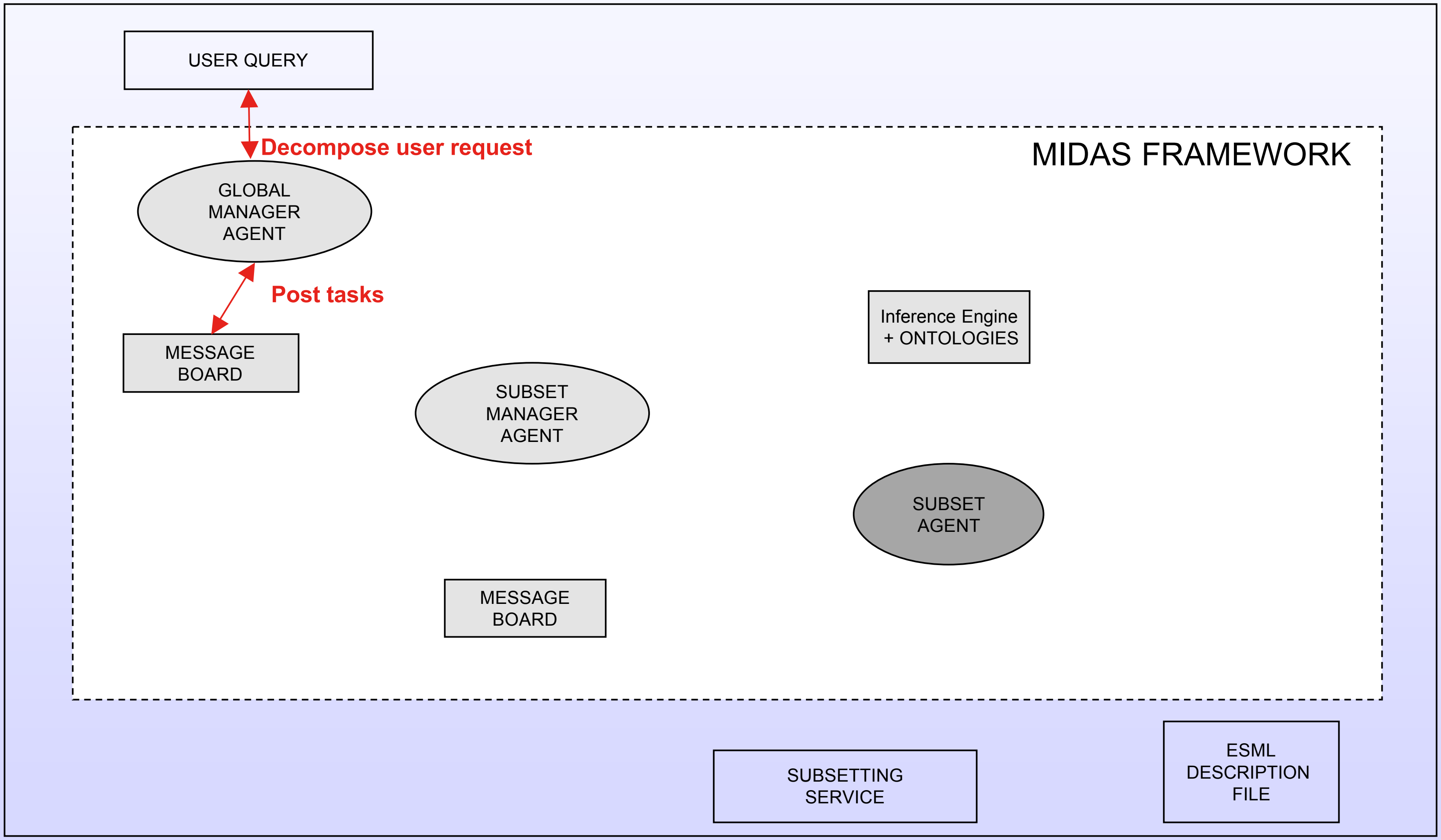
Advertise

SUBSET AGENT

The University of Alabama in Huntsville

June 22-25,2004

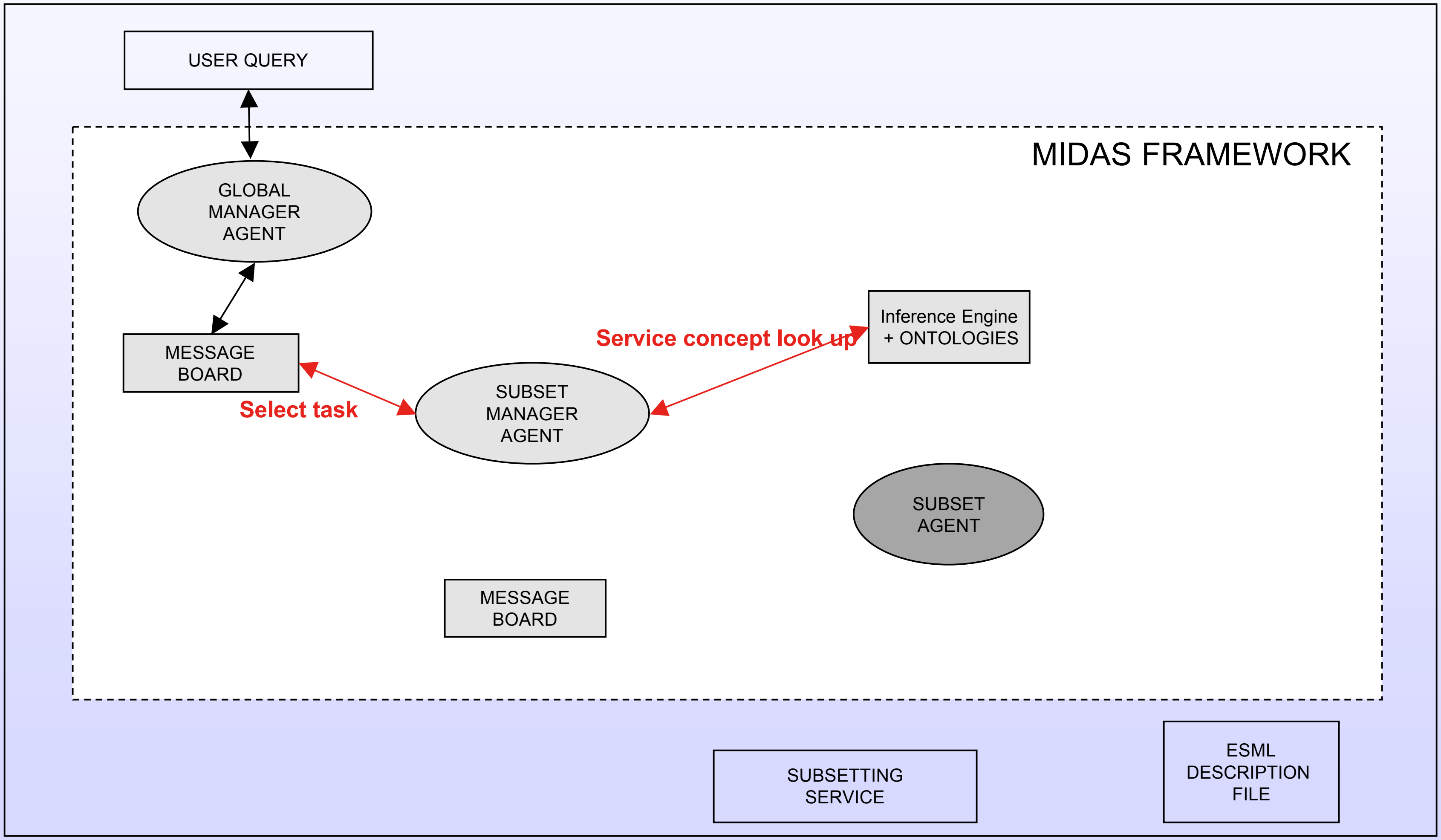
MIDAS in Action: Subset Query



The University of Alabama in Huntsville

June 22-25,2004

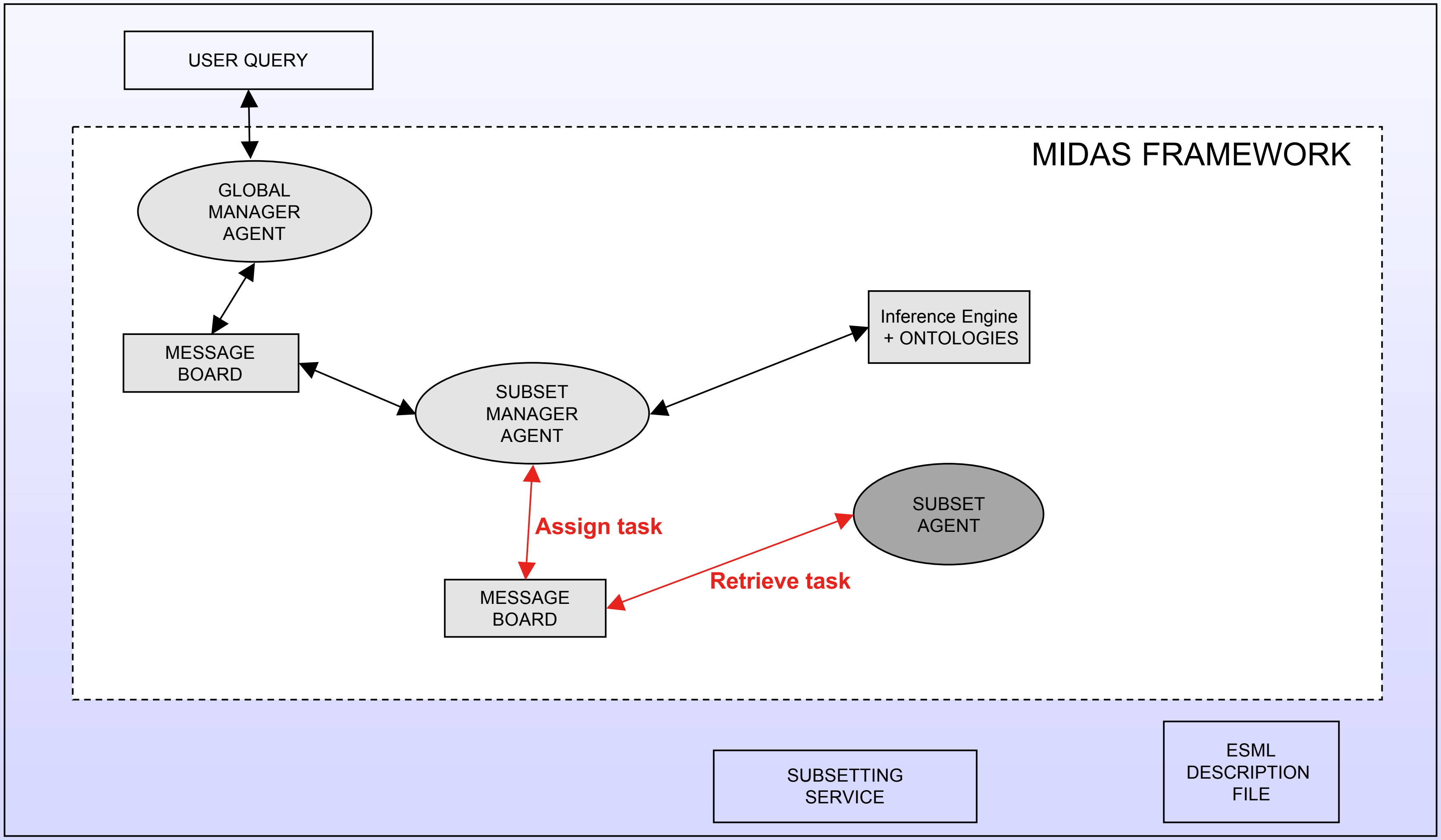
MIDAS in Action: Subset Query



UAH
The University of Alabama in Huntsville

June 22-25,2004

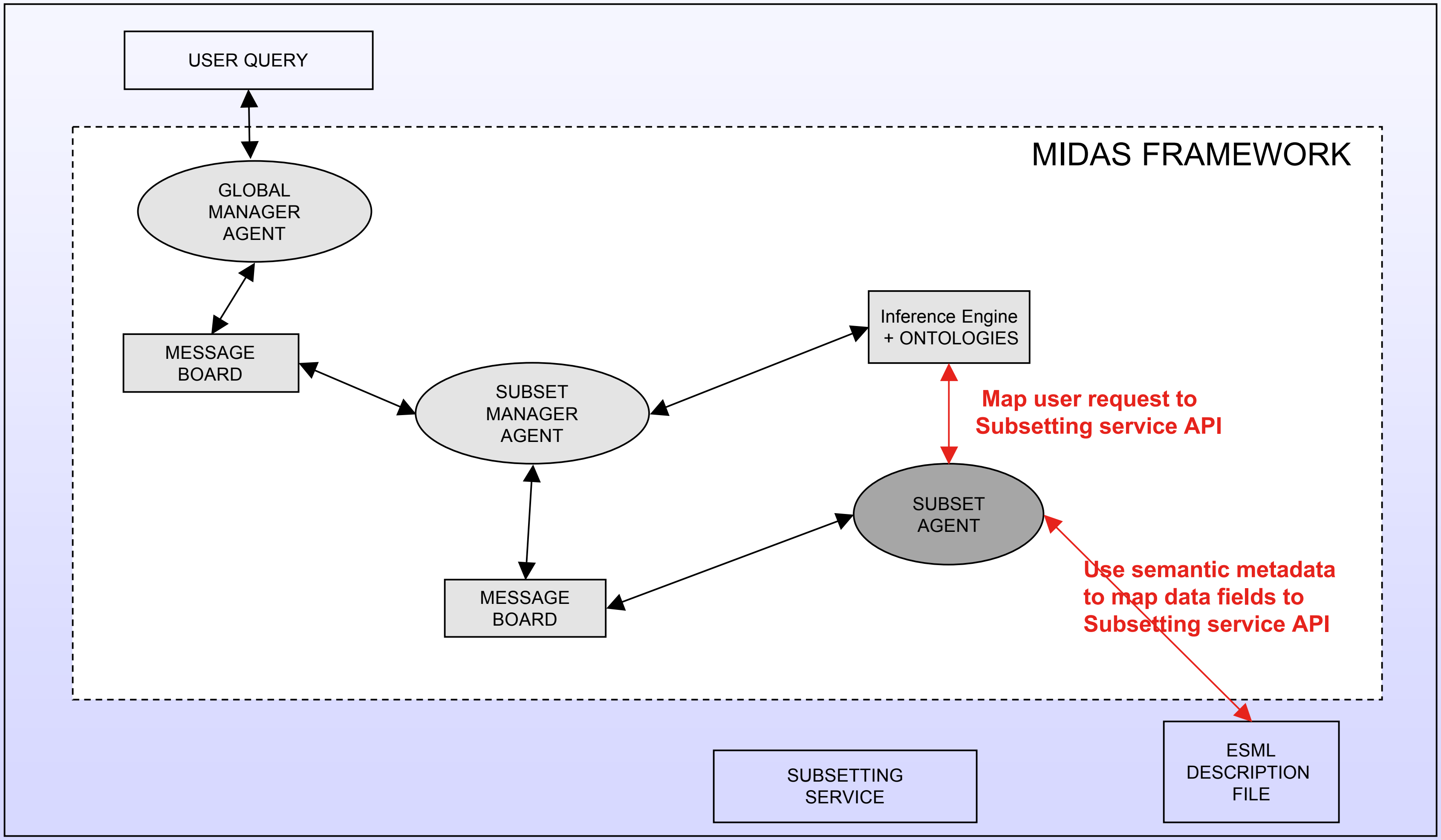
MIDAS in Action: Subset Query



UAH
The University of Alabama in Huntsville

June 22-25,2004

MIDAS in Action: Subset Query



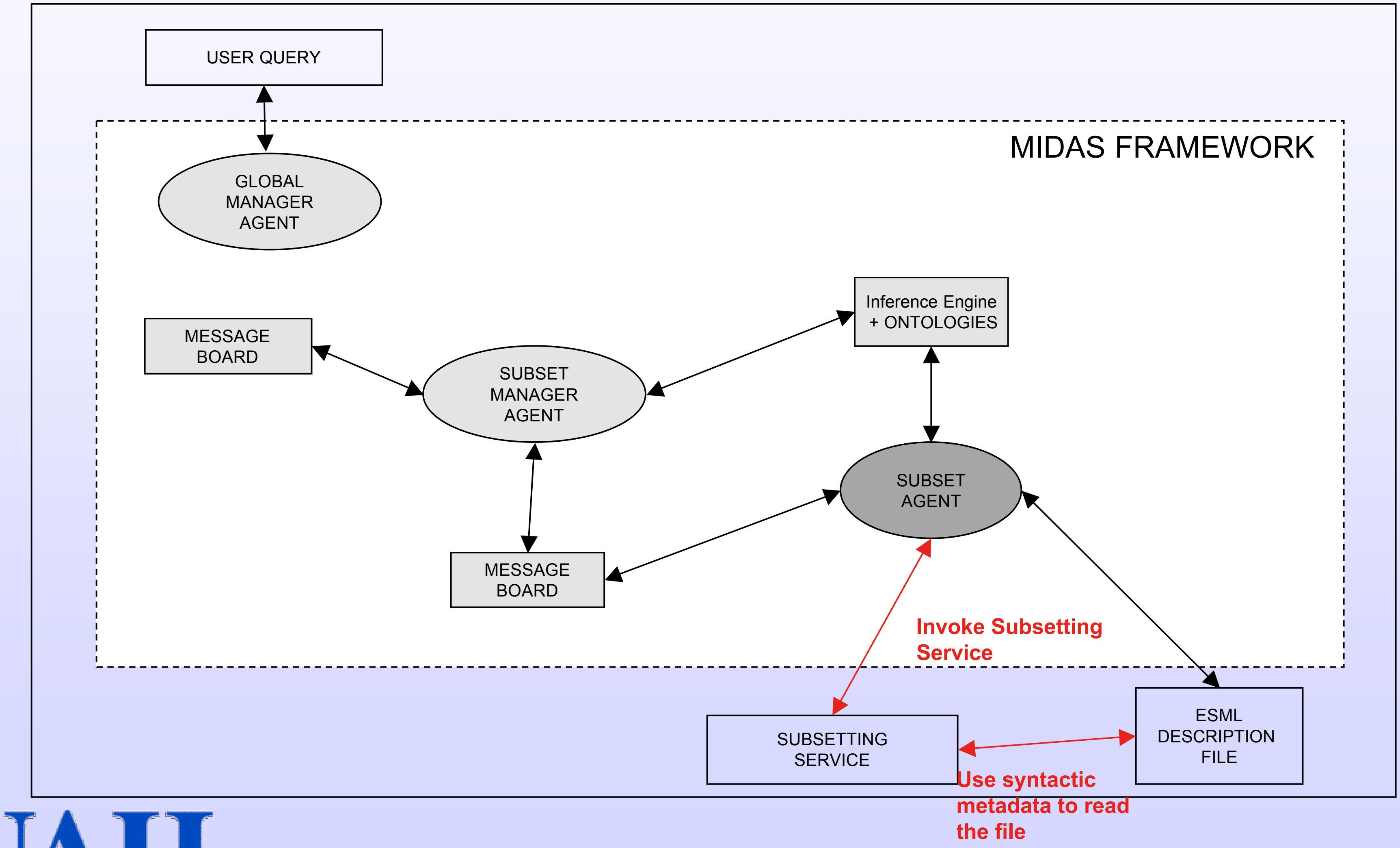
Map user request to Subsetting service API

Use semantic metadata to map data fields to Subsetting service API

The University of Alabama in Huntsville

June 22-25,2004

MIDAS in Action: Subset Query



June 22-25,2004

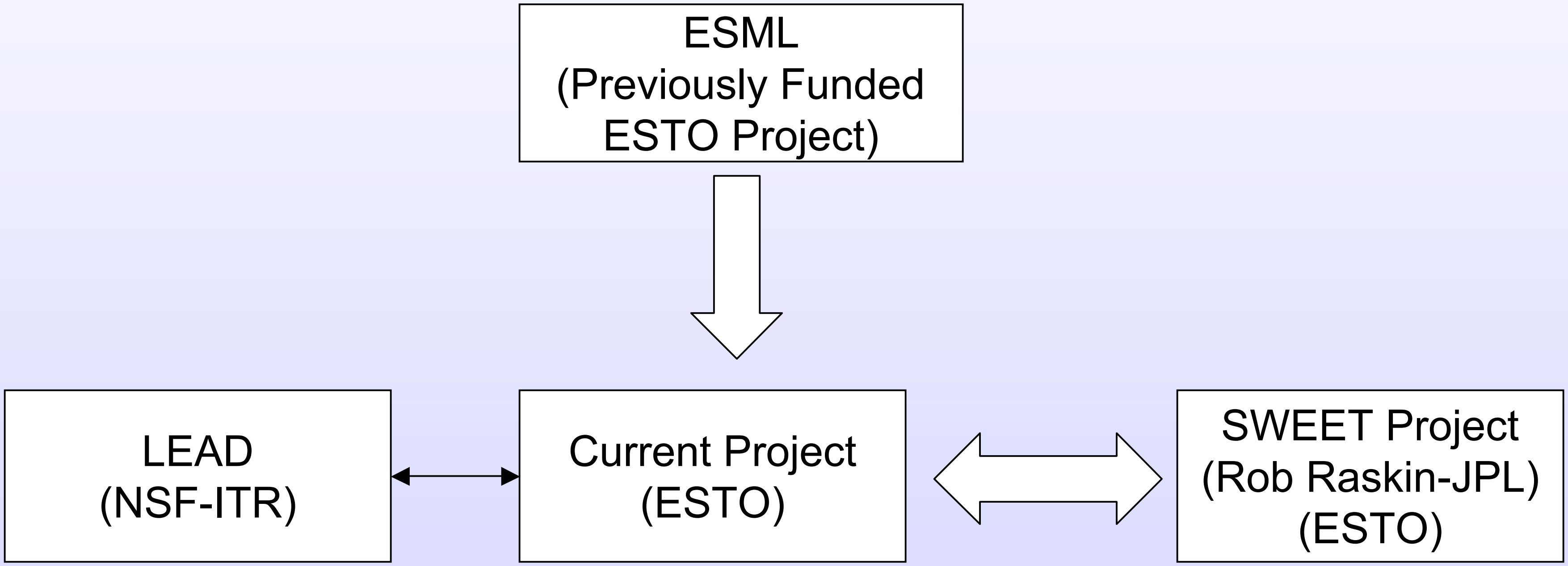
Demo: Simple Subsetting Query

- Given spatial and temporal bounds, subset a data file
  Domain manager selects a subset agent
  Subset agents performs the following actions:
  Navigate the data using ESML semantic metadata and the ontology
  Map the data to the subsetting service API

Publications/Presentations

- Ramachandran, R., H. Conover, S. Movva, and S. Graves, 2003: Using ESML in a Semantic Web Approach for Improved Earth Science Data Usability. *Semantic Web*, Sannibel, FL.
- Ramachandran, R., S. Movva, and S. Graves, 2003: Coupling Ontology with Earth Science Markup Language for Scientific Dataset Description. *Geological Society of America Meeting*, Seattle, WA.
- Movva, S., R. Ramachandran, X. Li, S. Khaire, K. Keiser, H. Conover, and S. Graves, Submitted 2004: Syntactic and Semantic Metadata Integration for Science Data Use. *Computers & Geosciences.*
- Ramachandran, R., S. Graves, S. Movva, and X. Li, 2004: Agent Framework for Intelligent Data Processing. *IEEE International Geoscience and Remote Sensing Symposium*, Anchorage, Alaska, IEEE.

Collaborations



ESML (Previously Funded ESTO Project)

LEAD (NSF-ITR)

Current Project (ESTO)

SWEET Project (Rob Raskin-JPL) (ESTO)

The University of Alabama in Huntsville

June 22-25,2004

Questions?